

汉语作为第二语言教学和 评估任务的工作分析^{*}

鹿士义^{1,2}

王二平²

¹ 北京大学对外汉语教育学院

² 中国科学院心理研究所

提要 本研究以问卷调查方式,通过对全球五大地区的 1135名被试对 64项任务评价的工作分析,确定了海外初级水平的汉语学习者和使用者重要的汉语使用任务,用描述性分析以及差异性检验分析了这些任务的重要性、使用频次以及难度等方面的差异。这些研究结果能为汉语教材的编写、考试模型的建立以及能力标准的研制提供重要的证据资料。

关键词 语言教学与测试 任务分析 CSL学习者和使用者

近十年来,表现评估(Performance Assessment)在语言教育领域得到了广泛关注和运用,在与第二语言习得和研究有关的语言测试与评估中,运用表现评估理论和方法研制的各种语言评估量表和测试,如 ASLPR(Australian Second Language Proficiency Ratings)、CEFR(Common European Framework of Reference for Languages Learning Teaching Assessment)以及美国外语教学学会(American Council on the Teaching of Foreign Languages ACTFL)的能力量表(Proficiency Guidelines)更是受到研究者和使用者的推崇,给第二语言的教学与测试带来了深刻变革,代表了语言教学研究领域的最新进展。

以表现评估为导向而研制的语言量表和测试大都采用行为定向的路子(action-oriented approach),即以语言学习者和使用者为中心,强调其能够使用语言做什么,能够完成什么样的交际任务,依据不同语言水平的人在各种条件和环境下完成语言活动或任务的行为表现来推测其语言能力。因此在量表的制定、教材的编写和测试的设计开发过程中,任务的选取以及任务分析至关重要(Bridgman & Carlson 1983, Hale, Tayör, Bridgman, Carson, Kroll & Kantor 1996, Bejar, Douglas, Jamieson, Nissan & Turner 2000, Butler, Eignor, Jones, McNamara & Sumi 2000, Cumming, Kantor, Powers, Santos & Tayör 2000, Enright, Grabe, Koda, Mosenthal, Mulcahy-Emt & Schede 2000, Jamieson, Jones, Kirsch, Mosenthal & Tayör 2000, Rosenfeld, Laung & Olman 2001)。确定这些任务以及任务的使用条件,属于工作分析(Job Analysis)的范畴。随着表现评估在教育领域中的广泛运用,工作分析在语言教

* 本项研究得到了国家汉办考试处、中国科学院心理研究所、北京大学对外汉语教育学院的大力支持和资助,谨此致以诚挚的谢意,同时还要感谢本研究在访谈、调查实施、编码、材料转写等过程中给予帮助的单位和个人以及匿名审稿专家的宝贵意见。

学与测试中也得到广泛的使用(王二平、谢小庆, 1994 Norris Brown, Hudson & Yoshioka, 1998 Long 2005)。《教育与心理测试标准》(American Educational Research Association et al, 1999)把工作分析作为测试内容效度的主要方法。Jane (1979)认为语言评估的第一步就是进行工作分析, McNamara (1996)认为保证测试最安全有效的方法就是着手进行工作分析。本项研究的主要目的,就是通过工作分析为将汉语作为第二语言的教学和测试确定需要用汉语完成什么样的任务,以此来确定初级阶段汉语教学和测试的相关知识范围和水平标准。

一 研究目的

本项研究旨在通过工作分析建立起汉语能力标准与具体任务之间的联系,将理论概念转化为有代表性的具体的语言使用任务,使其具有可操作性;通过工作分析确立的这些任务类型可以作为教材编写以及测试内容的直接来源,为考试提供内容效度的证据。

本项研究拟回答以下问题:对于初级汉语水平的学习者和使用者来说,哪些汉语使用任务在其生活和学习中最为重要?这些任务的使用频率和难度如何?这些任务的重要性、使用频次以及难度对亚洲、欧洲以及北美、大洋洲的初级汉语水平的学习者和使用者来说是否存在差异?在同一大洲中,学习者和使用者在目的语国家的语言环境中与其在自己国家语言环境中使用这些任务时,其重要性、使用频次以及难度上是否存在差异?

二 研究方法和程序

Brown (1995)认为,在进行语言需求分析时要求研究者必须考虑以下几个方面的因素:需求分析的对象、理论基础、信息特点、问题类型以及调查工具的类型等。国外为语言教学与测试所进行的语言使用任务的研究大多采用大规模问卷调查的方式进行。由于以汉语作为第二语言教学和评估对象的学习者与使用者,其地域分布广,而以汉语作为外语使用(CFL)的环境和以汉语作为第二语言使用(CSL)的环境则又有较大差异,学习者与使用者的背景各异,汉语水平参差不齐。要为不同水平的学习者和使用者确定所有应包括的语言使用任务以及相应的范围和水平,其难度是可以想见的。因此只有通过实证的工作分析,通过对汉语使用情况的实地考察,才能确定最合理的任务范围及最恰当的能力水平。本研究拟采用问卷调查的方式来进行研究。

为确定典型的汉语使用任务,本研究首先通过焦点小组访谈、个体深度访谈、专家干预等一系列方法和步骤形成初步的任务列表,然后将这些任务列表设计成调查问卷进行实地调查,最后对调查结果进行统计分析以获取典型的汉语使用任务。

2.1 分析工具

在如此大的范围内进行调查研究在国内外尚属首次,目前国内服务于第二语言教学的大规模语言使用任务的调查还不多见,可资利用的资源和分析工具并不多。已有的调查工具多数是对国外调查问卷的改良,难免会在“研究工具上出现文化差异”,因此本研究充分吸收了国外相关领域的研究成果,结合深度访谈等方法开发出《汉语使用任务调查问卷》作为此次研究的分析工具。

2.1.1 访谈与汉语使用任务的初步确定

访谈研究的目的在于初步了解汉语学习者学习和使用汉语的情况,收集和整理汉语使用任务,作为任务调查问卷的设计依据。

本研究的访谈规模定为 60 人,主要选择了北京、上海、广州、长春和大连等我国目前招收留学生人数较多且来自不同地域的留学生较集中的高校:北京大学(欧美)、第二外国语学院(韩国)、复旦大学(欧洲、大洋洲)、吉林大学(日本、俄罗斯、韩国)、东北财经大学(日本)、中山大学(泰国、印尼)和暨南大学(越南、马来西亚)等 7 所高校,考虑到学习者和使用者中不同的职业群体,又专门选择了上海外服公司和中智公司作为补充,以保证抽样的代表性。抽样层次主要有两个:

1) 地域:亚洲(东北亚、东南亚)、欧洲、北美洲、大洋洲

2) 汉语水平:初级水平、中级水平、高级水平

根据前期焦点小组座谈会形成的问题,采用关键事件法(Critical Incident Technique CIT)设计访谈提纲。2006 年 9 月先后在北京、上海、长春和广州等地采用了半结构化访谈方式进行个体深度访谈。

访谈时首先向被试说明研究目的,向被试出示访谈保密协议,在征得被试同意的基础上,对整个访谈过程进行录音,并将录音转写成文字资料,最终得到 37 份近 80 万字访谈记录。访谈结束后根据访谈纪录,将被试提及的有关汉语使用任务进行提取整理,形成编码本,邀请 9 位相关领域的专家,分成三个小组根据访谈记录和编码本,运用质性分析软件 NVivo 2.0 进行编码。三个组各自独立地对访谈记录进行编码,对编码的结果采用类目一致性(Category Agreement CA) $CA = 2 \times A / (B + A)$ 作为测量编码一致性指标(于涛,2004),采用损失比例减少(Proportional Reduction in Loss PRI^①)作为测量信度的指标(Rust & Coquil 1994)。三者的编码平均一致性系数 $CA = 0.879$, $PRI = 0.87$ 。最终整理出了汉语使用任务 126 项。这些汉语使用任务为设计调查问卷提供了第一手资料。

2.1.2 调查问卷的设计与开发

问卷共包括以下四个方面的内容:1)背景信息;2)在本国使用汉语的情况;3)在中国使用汉语的情况;4)汉语水平行为锚定量表。

通过访谈得到的语言使用任务,是问卷项目的主要来源,这些任务构成了问卷最初的任务条目。对照 CEFR(Common European Framework of Reference for Languages Learning Teaching Assessment)对语言使用任务的界定:“语言使用任务是在特定环境中人们使用语言实现某个特定目标或目的的活动”,我们对这些任务条目进行了修改与合并,然后用电子邮件将修改合并的任务条目发给海内外的汉语教学专家进行把关,以保证这些任务都是汉语学习者和使用者在目的语环境中所使用的最重要的任务。其次,在北京有关高校当面向教师和学生征询意见,最后将这些意见加以汇总,根据他们的意见对任务以及任务的描述进行修改,最终得到了 64 项任务列表。作为问卷条目的任务必须符合以下条件:1)任务必须描述得具体;2)任务应涉及到听说读写以及交互等不同的使用域(domain);3)任务要与教育、工作职业、个人和公众等环境背景相关。

为了尽可能多地调查语言使用任务,能把这些任务列表在问卷中尽可能多地罗列出来,

① $PRI = \frac{E(\max(L)) - E(L)}{E(\max(L))}$ $E(L)$ 是期望损失, $E(\max(L))$ 是期望最大损失。

考虑到问卷项目过多会影响被试的作答表现, 我们又把这些任务采用“打包”的方式, 用一个较为概括性的任务列出大任务, 然后再用举例的方式把相邻的任务加以合并。尽管整个问卷只有 64 项, 实际上所调查的语言使用任务要远远多于这个数目。

问卷主要从三个方面来对每一项任务进行评定:

1) 任务的重要性评定: 采用 Likert 五点量表。旨在确定每项任务的重要性。请被试根据自己日常生活、学习或工作中使用汉语的实际情况, 评定每一项活动的重要程度, 其中选项为“0”的表示在学习和工作中不需要这项任务。

2) 任务的使用频次评定: 采用 Likert 五点量表。请被试根据自己的实际情况回答这种活动的使用频率。

3) 任务的难易度评定: 采用 Likert 五点量表。请被试根据自己的判断回答做这种活动的难易程度。

2.1.3 行为锚定量表与汉语水平的确定

由于被试地域比较广, 各大洲的教育环境等因素的影响, 有时一次抽样的结果不能很好地满足统计分析的需要, 难以对被试的汉语水平加以判定。为了克服这些困难, 我们采用行为锚定量表评价法^② (Dessler 1994), 将汉语水平等级的叙述加以等级性量化, 将描述性关键事件评价法和量化等级评价法的优点结合起来。通过二次抽样来对被试的汉语水平等级进行评定。行为锚定量表的研制过程如下:

1) 找出听说读写四个方面中初级汉语水平、中级汉语水平和高级水平每个级别的关键事件, 请五位专家对这些关键事件进行评述, 断定这些任务对相应级别的被试来说是否能够很好地完成, 如中级汉语水平的被试所能完成的任务是初级汉语水平的被试所不能完成的, 高级汉语水平的被试所能完成的任务是中级汉语水平和初级水平所不能完成的。

2) 然后请另外五位专家对这些关键事件重新加以分配, 重新排序, 并用五点等级尺度评定法进行评定, 以判断它们是否有效地代表每个级别所要求的汉语水平。

3) 问卷在北京师范大学和对外经贸大学预测后, 我们把被试对这些“行为锚”的评价得分进行聚类分析, 并将聚类分析的结果请两位任课教师再次进行评定, 看看最终聚类的结果是否和被试的实际汉语水平相一致。两者之间的评价相关系数为 0.913。聚类出来的初级、中级、高级汉语水平也是与被试平时的汉语水平相符合的。

这样我们就可以用行为锚定量表, 通过聚类分析把被试的汉语水平有效地区分出初、中、高三级, 为下一步的统计分析奠定基础。

2.1.4 预测试

问卷形成后, 2006 年 11 月下旬, 在正式大规模施测前选择了北京师范大学和对外经贸大学各 30 名学生进行了问卷的预测。预测的目的, 一是检验问卷以及指导语表述得是否清楚; 二是检验问卷施测的程序是否合适。预测结束后主试与被试进行了座谈, 并对测试的

② 行为锚定等级评价法也称行为定位法, 行为锚定等级评价法是一种将同一职务工作可能发生的各种典型行为进行评分度量, 建立一个锚定评分表来对员工工作中的实际行为进行测评分级的评价方法。在同一个表现 (Performance) 维度中存在一系列的行为, 每种行为分别表示这一维度中的一种特定行为水平, 将行为水平按等级量化, 可以使评定的结果更有效, 更公平。

60 份问卷做了统计分析。经检验, 问卷主要统计指标如下: 问卷的 Cronbach alpha 信度为 0.845 分半信度 (Split half reliability) 为 0.78 内部一致性信度 (Internal consistency reliability) 为 0.87。

根据预测结果和反馈意见, 我们只对问卷的个别条目和指导语进行小的修改。问卷定稿后分别翻译成英、法、俄、日、韩、西、泰 7 种语言的版本, 并将各种语种的调查问卷制作成供网上调查的电子版, 同时将答卷制作成光电扫描版。

2.2 调查方式和抽样

限于经费和时间, 本研究的调查实施由国家汉语国际推广领导小组办公室委托当地的大学和汉语水平考试 (HSK) 考点来负责完成, 主要采用信函、电子邮件、网上作答等方式进行。调查对象的选择主要集中在东北亚的日本、韩国, 东南亚的新加坡、泰国和马来西亚, 北美的美国和加拿大, 欧洲的法国、俄罗斯、比利时、西班牙和意大利等国。调查从 2006 年 11 月开始至 2007 年 8 月结束, 共发放问卷 3000 多份。同时还在北京大学、北京师范大学、中国人民大学、北京外国语大学、首都师范大学、复旦大学、华东师范大学、大连外国语大学, 对利用暑期来华短期学习的汉语进修生进行了调查, 共发放问卷 600 多份。之所以选择这批短期学生进行调查, 主要是这部分学生刚来中国, 应该对在自己在本国汉语学习和使用的情况比较清楚, 在中国所受的“干扰”相对较小。

2.3 数据分析

2.3.1 聚类分析

主要用来分析被试在“汉语水平行为锚定量表”的作答, 用 Q 型聚类将所有的被试聚为初级、中级和高级三类不同的汉语水平。因为一般的教学模式和测试大都将考生按照不同的汉语水平分为三个级别, 并且有针对性地分别开设相应级别的课程、编写不同的配套教材。因此聚类分析都是依照初中高三个级别来进行的, 聚类分析为后面的统计分析奠定了基础。

2.3.2 描述性统计

1) 选项为 0 的百分比分析。“0”表示每项任务的重要性评定量表上“不需要这类活动”的意思。本研究认为超过一半以上的被试评定该项任务为“0”, 那么该项任务就被视为汉语学习者和使用者不需要这类活动。这项分析除了计算出每个级别中的评价总体的百分比以外, 还分别统计了该级别的亚洲、欧洲、美洲和大洋洲的各自评定的百分比。即使在总体评价中超过半数的被试认为不需要这类活动, 但不同洲的评价如果小于这个百分比的话, 那么这项任务也被视为所必需的任务。反之, 如果有一个大洲 50% 以上的被试将该项任务评定为“0”, 即使总体评价评定这个任务为重要, 该项任务也不应该视为典型的汉语使用任务。

2) 均值、标准差与标准误。计算出被试对每项任务的重要性 (除去“0”评定)、频次和难易度评定的平均值、标准差和标准误。本研究认为, 重要程度平均得分 3.0 以上、频次平均得分 3.0 以上、难易度得分 4 以下的任务为典型的汉语使用任务。

2.3.3 差异性分析

在每个水平上对任务的重要性、频次以及难易度分别进行洲与洲之间、国家与国家之间的差异性检验: 1) 采用 ANOVA 方法比较分析洲与洲之间的差异; 2) 采用配对 t 检验, 分别

比较三个洲中的国内与国外之间的差异；3)采用多层线性模型分析洲、国家、个体哪个是造成被试之间差异的最大来源。差异性检验可以用来判定这些任务在不同层次上一致性的程度和水平。

三 调查结果

本项研究共回收了 1179份调查问卷,其中有效问卷为 1135份,有效率为 96%。亚洲的被试人数为 697人,占总人数的 61.4%,欧洲为 175人,占 15.4%,美洲、大洋洲为 263人,占 23.1%。聚类结果显示,初级水平被试为 286人,占 25.1%,中级水平被试为 530人,占 46.6%,高级水平被试为 317人,占 27.9%。

被调查者中,汉语学习时间为 2—4年的占有相当大的比例,为 58.3%,而作为短期学习者学习时间在 1年以内的为 38.5%。有 72.5%的被试到过中国,多数为短期学习和旅游,为总调查人数的一半以上。多数人是作为留学生或旅游者来中国的,而来中国工作的也占有一定的比例,为 11.2%。被调查者最想提高的汉语能力主要是听和说的能力,其次为读的能力。写的能力和翻译能力基本持平,各占 10%左右。被调查者以后最想从事的与汉语有关系的工作是商业贸易和文化交流,而汉语教育和外交工作则仅为 7%左右。被试人口学特征统计详见表 1。

表 1 被试人口学特征统计

变量	类别	频次	百分比	缺失值(%)
性别	男	493	41.8	26(2.2)
	女	660	56.0	
年龄	15—20岁	300	25.4	36(3.1)
	20—30岁	532	45.1	
	30—40岁	163	13.8	
	40岁以上	148	12.6	
目前从事的工作	全职或部分工作	404	34.3	36(3.1)
	学生	638	54.1	
	军队服役	13	1.1	
	家庭主妇	32	2.7	
	其他	56	4.7	
汉语教育的最高水平	普通中学	170	14.4	6(0.5)
	大学预科	97	8.2	
	培训学校	143	12.1	
	大学本科	335	28.4	
	研究生	237	20.1	
	其他	191	16.2	

四 调查结果分析

4 1任务的必要性分析

如前所述,任务的必要性分析是指评价选项“0”在任务的重要性评价中所占的百分比。如果超过一半以上的被试评定该项任务为“0”的话,那么该项任务就被视为汉语学习者和使用者不需要这类活动。任务必要性分析的结果见附录。

分析结果表明,在本国情况的总体评价中,所调查的 20项汉语任务均为汉语使用任务所必需。任务 20“用汉语准备求职材料(如写求职信、写申请书、写个人简历)”,共有 28%的被试认为不需要这项任务,所占百分比为最高。对各大洲的本国情况分析表明,任务 11“在旅游场所用汉语帮助中国游客”(57%)、任务 12“为中国游客当导游”(53%)、任务 13“在商店帮助中国游客(如建议、付款、说明价格、帮助寻找失物)”(50%)、美洲、大洋洲中有一半或超过一半的被试认为这些任务不是典型的汉语使用任务。

对中国情况分析的结果表明,无论是总体评价还是各大洲的评价,所调查的 44项任务都是学习和工作中必要的任务。

4 2任务的重要性分析

任务的重要性分析是指通过描述性统计分析区分出哪项任务是重要的汉语使用任务,凡评定的平均值高于 3.0的(即该项任务在重要程度上为“重要”以上)则被视为汉语学习与使用中重要的汉语使用任务。剔除评价为“0”(不需要这类活动)的选项后,分别对各大洲被试的总体描述性评价结果和不同洲别的描述性评价结果进行统计及比较,结果见附录。

在总体评价中,平均值最高为 3.66 即任务 48“看病(如用汉语向医生描述症状)”,最低为 2.72 即任务 6“参加用汉语进行的正式面试或访谈(如应聘面试,接受新闻采访)”。64项任务中有 11项均值低于 3.0 占总数的 17%,其中本国情况 10项,中国情况 1项,见表 2。表 3列举了总体评价中的本国情况和中国情况均值得分前 5项的任务排序。

表 2 初级汉语水平均值低于 3.0的任务列表

类别	任务描述	均值
本国情况	4 用汉语参与课堂讨论(如讨论文化差异,讲述个人经历,谈论时事)	2.98
	17 用汉语接待中国人(如邀请、问候、请客吃饭)	2.98
	19 运用与娱乐有关的中文(如看中文电视节目、录像,看图书杂志、小说)	2.95
	12 为中国游客当导游(如用汉语指路、介绍有关景点、介绍注意事项、解答问题)	2.94
	7 理解与专业有关的学术或技术上的汉语用语/术语	2.89
	20 用汉语准备求职材料(如写求职信、写申请书、写个人简历)	2.89
	14 在机场帮助中国游客办理手续(如确认旅游团成员、登记检查、登机、领取行李)	2.85
	13 在商店帮助中国游客(如建议、付款、说明价格、帮助寻找失物)	2.81
	11 在旅游场所用汉语帮助中国游客(如填写表格、预订和选择旅游线路、安排日程、处理投诉)	2.76
	6 参加用汉语进行的正式面试或访谈(如应聘面试,接受新闻采访)	2.72
中国情况	52 使用图书馆(寻找所需图书,借还书)	2.94

表 3 初级汉语水平总体评价均值前 5 项的任务列表

类别	排序	任务描述	均值
本国情况	1	3 课后和同学用汉语交流（如一般的谈话, 讨论学习问题等）	3 26
	2	1. 用汉语处理课堂一般活动（如听老师讲课、提问、回答问题）	3 19
	3	5 与汉语教师 / 辅导员进行交流（如请假, 谈论与课程有关材料）	3 18
	4	15 在餐馆帮助中国游客就餐（如用汉语介绍菜单、点菜）	3 14
	5	9 为一般目的使用中文互联网（如浏览一般的中文网页, 用汉语聊天）	3 13
中国情况	1	48 看病（如用汉语向医生描述症状）	3 66
	2	36 乘火车 / 地铁 / 公共汽车（如用汉语购票、询问有关信息、查地图、查时间表）	3 57
		38 乘出租车（如用汉语指路、问价）	3 57
		42 阅读专业的中文资料、论文	3 57
	3	33 与同事朋友用汉语进行社交活动（如喝酒聊天, 去卡拉 OK 接受 / 拒绝邀请）	3 54
	4	26 为一般目的使用电话（如用汉语与同学、朋友、老师、同事等联系）	3 51
	5	21. 用汉语处理课堂一般活动（如听老师讲课、提问、回答问题）	3 5
		34 与上司和客户在工作中用汉语交流（如要求 / 提供帮助, 获取信息）	3 5

从对不同大洲的总体评价分析可以看出, 均值最高分为 4 17 (欧洲), 即任务 36“乘火车 / 地铁 / 公共汽车 (如用汉语购票、询问有关信息、查地图、查时间表)”, 最低分为 2 5 (美洲、大洋洲), 即任务 17“用汉语接待中国人 (如邀请、问候、请客吃饭)”, 有 26 项任务的评价均值低于 3 0。表 4 列举了不同大洲总体评价中平均均值为 3 5 分以上的任务排序。

表 4 初级汉语水平各大洲总体评价平均均值为 3 5 以上的任务列表

类别	排序	任务描述	均值
本国情况	1	1. 用汉语处理课堂一般活动（如听老师讲课、提问、回答问题）	3 54
中国情况	1	36 乘火车 / 地铁 / 公共汽车（如用汉语购票、询问有关信息、查地图、查时间表）	3 81
	2	38 乘出租车（如用汉语指路、问价）	3 78
	3	33 与同事朋友用汉语进行社交活动（如喝酒聊天, 去卡拉 OK 接受 / 拒绝邀请）	3 69
		35 处理在机场的有关问题（如用汉语询问行李、回答报关问题）	3 69
	4	48 看病（如用汉语向医生描述症状）	3 66
	5	39 观光旅游（如收集信息、确定旅行目的地）	3 65
		26 为一般目的使用电话（如用汉语与同学、朋友、老师、同事等联系）	3 51

对洲与洲之间重要性的评定结果, 用 ANOVA 方法分析、比较洲与洲之间的差异。结果显示, 洲与洲之间对任务重要性的评价, 无论是总体评价、对本国情况的评价还是对中国情况的评价差异均不显著 ($F_{总} = 7.58$ $P_{总} = 0.470$, $P > 0.001$, $F_{本} = 2.412$ $P_{本} = 0.92$ $P > 0.001$, $F_{中} = 2.246$ $P_{中} = 1.08$ $P > 0.001$)。采用配对 t 检验, 分别比较三大洲中的本国情况与中国情况之间的差异, 结果见表 5。分析结果表明, 各大洲中对任务重要性的评定, 本国情况与中国情况均有显著差异。这个结果说明我们在教学活动或教材编写以及进行教学评估时, 选取任务时要特别注意语言环境的差异性。

表 5 初级汉语水平三大洲中本国情况配对 检验

洲别	比较类目	国内 (A)		国外 (B)		差值 (A-B)	P值
		均值	样本数	均值	样本数		
亚洲	难度	3.643	137	3.710	137	-0.067	0.170
	重要性	2.299	139	2.724	139	-0.424	0.000
欧洲	难度	3.660	31	3.960	31	-0.300	0.005
	重要性	1.859	33	3.072	33	-1.213	0.000
美洲、大洋洲	难度	3.244	24	3.578	24	-0.335	0.234
	重要性	1.971	28	3.051	28	-1.080	0.001

注: P值为双尾检验结果。

为了进一步探讨洲、国家、个体哪个是造成被试评价的最大的来源,我们将被试的评定分为个体、国家、洲三层,采用多层线性模型分析了这些差异的来源,表 6显示,在初级水平的被试中,D_1_64、F_1_20和 D_21_64受到国家之间差异的影响,大洲之间的差异不产生影响,个体之间的差异均产生显著影响。因此,可以认为初级水平个体之间的差异主要受到国家和个体差异的影响。

表 6 初级汉语水平多层线性模型分析结果

固定效应	D_1_64	I_1_64	D_1_20	F_1_20	I_1_20	D_21_64	I_21_64
	N=227	N=232	N=220	N=219	N=226	N=199	N=206
常数项	3.493** (0.212)	2.649*** (0.081)	3.377** (0.196)	2.003* (0.216)	2.057** (0.137)	3.664** (0.185)	2.958** (0.154)
随机效应	D_1_64	I_1_64	D_1_20	F_1_20	I_1_20	D_21_64	I_21_64
	N=227	N=232	N=220	N=219	N=226	N=199	N=206
常数项	0.000	0.000	0.000	0.000	0.000	0.000	0.040
(大洲)	—	—	—	—	—	—	(0.066)
常数项	0.438* (0.261)	0.000	0.302 (0.202)	0.514** (0.289)	0.060 (0.060)	0.290* (0.172)	0.000
(国家)		—					—
残差项	0.838*** (0.081)	1.511*** (0.141)	1.121*** (0.110)	0.461*** (0.046)	1.608*** (0.153)	0.819*** (0.085)	1.620*** (0.161)

注: ***表示 $P < 0.001$; **表示 $P < 0.01$; *表示 $P < 0.05$ 。I=重要性, D=难度, F=频次

五 讨论与启示

通过问卷调查,我们总结出了初级汉语水平的被试使用的典型任务(重要性 ≥ 3 且难度 ≤ 4)共有 80项,其中本国情况 20项,中国情况 60项。在这些典型任务中,既有以一种交流方式为主的任务,也有涉及两种交流方式的任务,结果见表 7。这些典型任务口头交流占 54%,书面语交流占 46%。学习者和使用者实际运用和使用的汉语传统的听说读写四种交流方式并不是孤立存在的,以两种交流方式为主的任务占有相当大的比例。传统的语言教学和测试不注重真实的语言使用情况,习惯于将听、说、读、写分开来进行。因此在教材编写

和课程设置以及构造测试任务时应参照真实使用各种任务的比例来进行。不仅应该包括传统的分离式的语言教学任务项目,还应该包括听说、读写等综合性的语言教学任务项目。在设计试题时不仅要有测试听说读写单一的语言能力的试题,同时还应该有兼顾测试综合语言能力的试题。

在这些典型任务中口头交流的比例高于书面语交流,这是汉语作为第二语言教学的一种特点。背景信息调查也显示出,在应试最想提高的汉语能力方面,听和说的口头交际能力所占比例为最高,分别为 32.8%、31.6%,而读和写的书面语交际能力所占比例分别为 13.9%和 10.6%,远远小于口头交际能力。

表 7 初级汉语水平典型任务的交流方式

分类	交流方式	任务数量	任务描述举例
本国情况	一种交流方式为主	听	2 听老师讲课
		说	6 用汉语在课堂上进行演讲或展示
		读	3 浏览一般的中文网页
		写	3 用汉语写私人信函
		小计	14
	两种交流方式为主	听说	4 为一般目的使用电话
		读写	2 用汉语回复电子邮件
		小计	6
中国情况	一种交流方式为主	听	1 听老师讲课
		说	12 用汉语问路
		读	15 乘火车 /地铁 /公共汽车 (查地图、时间表)
		写	8 使用手机给同事或朋友发送、回复短信
		小计	36
	两种交流方式为主	听说	20 为一般目的使用电话
		读写	4 处理传真、电子邮件
		小计	24

语言使用任务的核心内容是语言输入与输出的内容,因此输入与输出之间的关系是任务的重要特点,也是给任务分类的重要依据。Bachman & Palmer (1996)认为测试任务分类的重要标准是输入与所期待反应 (expected response)之间的关系。在初级阶段典型的汉语使用任务中,产出型任务所占比例最高,为 38%,其次为交互性任务,占 36%,接受型任务再次之,占 26%。语言活动的实质就是接收、产出和交换各种类型的文本 (text),因此它是语言使用任务的核心。“文本”在接受型任务中是语言输入的内容,在产出型任务中是语言输出的内容,在互动型任务中既是输入也是输出。“文本”信息是教材编写、设计考试时选择任务内容的参考,即在教学和评估中为考生设置什么内容的语言输入,以及预期考生有什么内容的语言输出。以下归纳了初级水平典型任务的文本类型,其中口头文本计 13 种,书面文本 36 种。这些文本可以作为教材编写、测试初级任务内容的直接来源,具体如下:

口头文本: 1) 本国情况: 发问、答问、演讲、谈话、讨论; 2) 中国情况: 讨论、发问、答问、演讲、谈话、谈论、交流、询问。

书面语文本: 1) 本国情况: 请假条、电子邮件、网页、菜单、私人信函、明信片、便条; 2) 中国情况: 请假条、手机短信、一般电子邮件、传真、商业信函、一般中文网页、专业中文网页、地图、火车时刻表、站牌、宾馆表格、地图、旅游介绍、车票、银行账单(表格)、会员卡、报纸、杂志、广告、产品说明书、购物单、商品说明、菜单、贺卡、讲话稿、信封、明信片、书单等。

调查结果显示, 汉语作为第二语言与汉语作为外语的学习和使用过程中具有明显的差异。在同一大洲中, 初级阶段的学习者在自己本国的语言环境中使用汉语与在目的语国(中国)的语言环境中使用汉语具有明显的不同。因此在教材编写以及考试的设计上要注意这种差异性。不能把具有很强地域性的语言任务表现出来, 从另外一个方面来说, 要在语料上尽量避免地域性强的任务, 否则就会失去语料的真实性。尤其是对测试而言, 选择真实性高的任务一般分为两个步骤: 第一, 辨别出目的语使用域的任务特征, 并依此建构任务特征的框架; 第二, 选择或设计具有这些特征的样本作为测试任务。交际测试其实是间接的行为测试, 所以测试的真实性取决于语言测试行为与非测试语言行为的一致性。Bachman & Palmer(1996)认为对真实性的定义应扩展为两个方面: 1) 目的语(外语)使用任务及情景的特征与测试任务及情景的特征应保持一致; 2) 语言使用者特征与受试者特征应保持一致。其中第二方面尤其重要, 因为该方面涉及到的语言能力才是测试的真正对象。

参考文献

- 王二平 谢小庆 (1994) 银行保险业经济系列专业职称资格的职务分析, 《心理学动态》第 1 期。
- 于广涛 (2004) 某石油化工厂安全文化的多元方法研究, 中国科学院研究生院博士学位论文。
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999) *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.
- Bachman, L. F. & A. S. Palmer (1996) *Language Testing in Practice*. Oxford: Oxford University Press.
- Bejar, I., D. Douglas, J. Janieson, S. Nassar & J. Tumer (2000) *TOEFL 2000 Listening Framework: A Working Paper* (TOEFL Monograph Series Report No. 19). Princeton, NJ: Educational Testing Service.
- Bridgeman, B., & S. Carlson (1983) *Survey of Academic Writing Tasks Required of Graduate and Undergraduate Foreign Students* (TOEFL Research Report No. 15). Princeton, NJ: Educational Testing Service.
- Brown, J. D. (1995) *The Elements of Language Curriculum: A Systematic Approach to Program Development*. New York: Heinle & Heinle.
- Butler, F. A., D. Eignor, S. Jones, T. McNamara & B. Sumi (2000) *TOEFL 2000 Speaking Framework: A Working Paper* (TOEFL Monograph Series Report No. 20). Princeton, NJ: Educational Testing Service.
- Cumming, A., R. Kantor, D. Powers, T. Santos & C. Taylor (2000) *TOEFL 2000 Writing Framework: A Working Paper* (TOEFL Monograph Series Report No. 18). Princeton, NJ: Educational Testing Service.
- Dessler, G. (1994) *Human Resource Management* (6th ed.). 北京: 中国人民大学出版社, 1996.
- Enright, M., W. Grabe, K. Koda, P. Mosenhall, P. Mukkahy-Emt & M. Schedle (2000) *TOEFL 2000 Reading Framework: A Working Paper* (TOEFL Monograph Series Report No. 17). Princeton, NJ: Educational Testing Service.
- Hale, G., C. Taylor, B. Bridgeman, J. Carlson, B. Kroll & R. Kantor (1996) *A Study of Writing Tasks Assigned in Academic Degree Programs* (TOEFL Research Report No. 54). Princeton, NJ: Educational Testing Service.

Jamieson J, S Jones J Kirsch P Mosenthal C Taylor (2000) TOEFL 2000 Framework A Working Paper (TOEFL Monograph Series Report No. 16). Princeton NJ: Educational Testing Service

Jones R L (1979) Performance testing of second language proficiency. In E J Friere & F B Hinojosa (eds.), Concepts in Language Testing: Some Recent Studies 50—57. Washington DC: TESOL

Long M (2005) Second Language Needs Analysis. Cambridge: Cambridge University Press

McNamara T F (1996) Measuring Second Language Performance. London and New York: Longman

Norris J M, J D Brown T Hudson & J Yoshioka (1998) Designing Second Language Performance Assessment. Honolulu: University of Hawai'i Press

Rosenfeld M, S Laung & P K Olman (2001) The Reading, Writing, Speaking, and Listening Task: Important for Academic Success at the Undergraduate and Graduate Levels. TOEFL Monograph Series Report No. 21. Princeton NJ: Educational Testing Service

Rust R T & B Cooil (1994) Reliability measures for qualitative data: theory and implications. Journal of Marketing Research 31 1, 1—14

附录 初级水平汉语使用任务必要性评价与重要性、频次和难度评价一览表

任务描述	总体				亚洲			欧洲			美洲、大洋洲		
	N= 194				N= 48			N= 29			N= 28		
	I(M)	SD	SE	0%	I(M)	F(M)	D(M)	I(M)	F(M)	D(M)	I(M)	F(M)	D(M)
第一部分 本国情况													
1 用汉语处理课堂一般活动(如听老师讲课、提问、回答问题)	3.19	1.27	0.1	12	3.12	2.04	3.47	3.61	3.18	2.8	3.27	2.76	2.5
2 用汉语在课堂上进行演讲或展示(如告诉同学怎么做某事或使用某物)	3.09	1.27	0.1	13	3.07	1.62	3.89	3.39	2.65	2.7	3	2.71	3
3 课后和同学用汉语交流(如一般的谈话,讨论学习问题等)	3.26	1.26	0.1	14	3.25	1.85	3.47	3.08	1.71	3.2	3.42	2.16	3.2
4 用汉语参与课堂讨论(如讨论文化差异,讲述个人经历,谈论时事)	2.98	1.2	0.09	16	2.94	1.58	4.01	3.27	2.06	3.4	3.05	2.38	3.2
5 与汉语教师/辅导员进行交流(如请假,谈论与课程有关材料)	3.18	1.24	0.1	14	3.14	1.86	3.52	3.11	2.47	3.2	3.43	2.5	3.1
6 参加用汉语进行的正式面试或访谈(如应聘面试,接受新闻采访)	2.72	1.31	0.11	24	2.54	1.28	4.12	3.3	1.18	3.8	3.69	1.68	3.3
7 理解与专业有关的学术或技术上的汉语用语/术语	2.89	1.35	0.11	19	2.8	1.4	4.03	3.08	1.53	3.8	3.32	1.79	3.2
8 处理汉语的电子邮件(如阅读汉语电子邮件,用汉语回复)	3.01	1.13	0.09	17	2.97	1.83	3.54	2.9	1.29	3.9	3.44	1.56	3.1
9 为一般目的使用中文互联网(如浏览一般的中文网页,用汉语聊天)	3.13	1.18	0.09	16	3.13	2.28	3.42	2.9	1.12	4.1	3.29	1.8	3.2
10 为学习或者工作目的使用中文互联网(如寻找资料,浏览专业网页)	3.08	1.15	0.09	19	3.07	1.49	3.49	3.1	1.47	3.9	3.12	1.75	3.4
11 在旅游场所用汉语帮助中国游客(如填写表格、预订和选择旅游线路,安排日程,处理投诉)	2.76	1.28	0.11	25	2.66	1.34	3.82	3.44	1.29	3.6	3.27	1.29	3.3

12 为中国游客当导游(如用汉语指路、介绍有关景点、介绍注意事项、解答问题)	2 94	1 25	0 1	19	2 87	1 45	3 48	3 18	1. 35	3 5	3. 42	1. 38	3 4
13 在商店帮助中国游客(如建议、付款、说明价格、帮助寻找失物)	2 81	1 18	0 1	20	2 8	1 36	3 57	2 75	1. 47	3 3	3	1. 29	3 2
14 在机场帮助中国游客办理手续(如确认旅游团成员、登记检查、登机、领取行李)	2 85	1 18	0 09	18	2 84	1 5	3 41	2 58	1. 18	3 5	3. 13	1 5	3 5
15 在餐馆帮助中国游客就餐(如用汉语介绍菜单、点菜)	3 14	1 21	0 1	18	3 16	1 31	3 96	2 91	1. 12	3	3 2	1. 42	3 1
16 在突发事件中为中国人提供帮助(如病症、伤害、事故、报告罪行)	3 02	1 18	0 09	18	2 95	1 47	3 74	3 18	1. 24	3 7	3 5	1. 42	3 4
17 用汉语接待中国人(如邀请、问候、请客吃饭)	2 98	1 21	0 1	16	3 06	1 71	3 47	2 92	1. 35	3 3	2 5	1. 63	3 3
18 与亲戚朋友联络(如用汉语写私人信函、明信片、便条)	3 07	1 26	0 1	12	3 1	2 32	3 46	2 86	1. 65	3 3	3	1. 42	3 4
19 运用与娱乐有关的中文(如看中文电视节目/录像,看图书杂志、小说)	2 95	1 27	0 11	25	2 95	1 42	4 11	2 83	1. 35	3 9	3	1. 58	3 5
20 用汉语准备求职材料(如写求职信、写申请书、写个人简历)	2 89	1 26	0 11	28	2 86	1 38	4 08	3	1. 29	3 8	3	1. 38	3 3
第二部分 中国情况													
21 用汉语处理课堂一般活动(如听老师讲课、提问、回答问题)	3 5	1 3	0 1	9	3 39		3 73	4		3 5	3. 63		3 4
22 用汉语在课堂上进行演讲或展示(如告诉同学怎么做某事或使用某物)	3 41	1 26	0 1	10	3 34		3 95	3 78		3 6	3. 46		3 7
23 课后和同学用汉语交流(如一般的谈话,讨论学习问题等)	3 35	1 32	0 1	9	3 38		3 7	3 24		3 4	3. 31		3 7
24 用汉语参与课堂讨论(如讨论文化差异,讲述个人经历,谈论时事)	3 27	1 22	0 09	8	3 22		3 97	3 22		3 7	3. 54		3 9
25 与汉语教师/辅导员进行交流(如请假、谈论与课程有关的材料)	3 28	1 22	0 09	7	3 23		3 7	3 47		3 5	3. 32		3 6
26 为一般目的使用电话(如用汉语与同学、朋友、老师、同事等联系)	3 51	1 23	0 09	6	3 51		3 76	3 53		3 8	3 5		3 9
27 因业务需要而使用电话(如与上司或下属、业务伙伴、客户联系)	3 49	1 2	0 1	11	3 54		3 97	3 18		3 9	3. 46		3 9
28 使用手机用汉语给同事或朋友发送、回复短信	3 49	1 21	0 09	7	3 48		3 59	3 61		4	3. 46		3 6
29 处理一般的中文电子邮件(如阅读和回复同学、朋友或老师的电子邮件)	3 37	1 3	0 1	8	3 38		3 65	3 28		4	3. 38		3 8
30 处理业务上的中文留言(如传真、电子邮件、商业信函等)	3 41	1 18	0 1	13	3 38		3 77	3 24		4 1	3. 73		3 7
31 为一般目的使用中文互联网(如浏览一般的中文网页、用汉语聊天)	3 21	1 34	0 1	9	3 16		3 72	3		4	3. 58		3 9

32 为学习或工作目的使用中文互联网 (如寻找资料、登录专业的网页)	3 39	1 21	0 1	9	3 28		3 64	3 38		4 1	3 92		4
33 与同事朋友用汉语进行社交活动 (如喝酒聊天、去卡拉 OK 接受 / 拒绝邀请)	3 54	1 27	0 1	5	3 45		3 48	3 63		4 1	4		3 7
34 与上司和客户在工作中用汉语交流 (如要求 / 提供帮助、获取信息)	3 5	1 19	0 1	12	3 39		3 93	3 47		3 9	4		3 8
35 处理在机场的有关问题 (如用汉语询问行李、回答报关问题)	3 48	1 18	0 09	9	3 34		3 61	3 76		3 9	3 96		3 7
36 乘火车 / 地铁 / 公共汽车 (如用汉语购票、询问有关信息、查地图、查时间表)	3 57	1 15	0 09	7	3 41		3 37	4 17		3 7	3 86		3 8
37 旅馆住宿 (如用汉语预订宾馆登记、处理有关设备及服务问题)	3 49	1 15	0 09	6	3 32		3 23	3 94		3 8	3 93		3 7
38 乘出租车 (如用汉语指路、问价)	3 57	1 19	0 09	3	3 43		3 28	4 05		3 4	3 86		3 6
39 观光旅游 (如收集信息、确定旅行目的地)	3 46	1 23	0 1	6	3 34		3 18	3 89		3 5	3 71		3 6
40 看电影、欣赏戏剧、参与俱乐部等娱乐活动 (如预订和购票、办理会员卡)	3 17	1 29	0 1	9	2 95		3 49	3 56		3 8	3 88		3 5
41 阅读一般的中文报纸、杂志	3 26	1 33	0 1	7	3 17		3 8	3 28		4 1	3 72		4
42 阅读专业的中文资料、论文	3 57	2 96	0 25	16	3 16		4 2	5 59		4	3 84		3 7
43 阅读中文的广告及产品说明书	3 01	1 32	0 1	7	2 87		3 69	3 41		4	3 41		3 9
44 去商店或超市用汉语购买日用品 (如生活用品、服装鞋帽、礼品等)	3 27	1 28	0 1	5	3 21		3 3	3 39		3 7	3 46		3 6
45 用汉语买大件商品 (如冰箱和电视、长期生活所需商品)	3 09	1 33	0 11	11	3 01		3 6	3 31		3 8	3 39		3 7
46 在饭店 / 快餐店就餐 (如用汉语点菜、订餐、叫外卖)	3 27	1 21	0 09	5	3 21		3 2	3 17		3 8	3 62		3 6
47 在路上与陌生人交流 (如用汉语问路)	3 28	1 22	0 09	7	3 17		3 39	3 37		3 6	3 71		3 8
48 看病 (如用汉语向医生描述症状)	3 66	1 13	0 09	7	3 65		3 76	3 56		4 2	3 78		3 8
49 在银行使用汉语 (如用汉语开户、换钱、存 / 取款、转账)	3 26	1 32	0 1	9	3 17		3 7	3 39		4 2	3 62		3 6
50 在邮局使用汉语 (如邮寄信件和明信片)	3 33	1 2	0 09	8	3 3		3 46	3 26		4	3 48		3 6
51 在书店用汉语买书 (寻找所需图书、个人预订)	3 17	1 24	0 1	8	3 08		3 26	3 47		3 7	3 36		3 6
52 使用图书馆 (寻找所需图书、借还书)	2 94	1 27	0 1	9	2 77		3 35	3 28		4 2	3 46		3 5
53 租住房屋 (如用汉语寻找公寓与房东、中介机构交流)	3 18	1 31	0 11	13	3 08		3 8	3 29		4 3	3 57		3 4
54 用汉语处理与出入境管理局、公安局、使馆等官方机构有关的问题 (如申请护照延期、民事纠纷、交通违章)	3 45	1 31	0 11	12	3 45		4 09	3 11		4 2	3 69		3 8

55 出席婚礼、葬礼、毕业典礼等各种正式场合(如送礼、写贺卡、发表讲话)	3 1	1 28	0 1	10	3 01		3 6	3 33		3 9	3 35		3 8
56 参加中国的传统节日活动(如春节、中秋节等)	3 14	1 24	0 1	6	3 09		3 4	3 17		4	3 33		3 7
57 用中文准备个人介绍和求职材料(如写求职信、写申请信、写个人简历)	3 14	1 33	0 11	12	3 04		4 02	3 16		4 3	3 54		3 8
58 用汉语参与进行的正式面谈(如参与应聘面试、接受新闻采访)	3 01	1 39	0 11	12	2 78		4 04	3 29		4	3 84		3 7
59 用中文处理日常工作的常规事务(如安排会议、签到)	3 15	1 21	0 1	11	2 97		3 84	3 47		4 3	3 79		3 4
60 用中文参与商务贸易,进行商业谈判	3 25	1 37	0 11	12	3 13		4 19	3 44		3 9	3 7		3 5
61 用中文参与工作会议讨论(如讨论成本、收益、未来发展)	3 1	1 42	0 11	12	2 86		4 13	3 56		4 2	3 92		3 8
62 用中文在工作会议上做报告/展示	3 24	1 37	0 11	15	3 03		4 14	3 65		4 2	3 95		3 6
63 阅读中文项目报告、学术论文	3 22	1 41	0 12	15	3 01		4 27	3 24		4 1	4 08		3 8
64 撰写中文评估意见/摘要、学术论文	3 24	1 45	0 13	17	3 04		4 29	3 19		4 2	4 09		3 8

A Working Analysis of the Tasks of Teaching and Assessment of Chinese as a Second Language

LIU ShiYi and WANG ErPing

Abstract This study aims to aid in translating the theoretical framework into task statements that overseas learners and users of Chinese need to perform in order to fulfill their communication tasks. To this end, 1135 CSL learners and users from Asia, Europe, Americas, and Oceania were chosen to rate 64 task statements in the importance, frequency and difficulty levels to successful communication. Responses were analyzed by respondent groups as well as specific sub-groups. These findings support the use of this pool of tasks in the preparation of both teaching materials and assessment measures for CSL learners and users across a wide range of areas.

Key Words language teaching and testing; task analysis; CSL learners and users

作者简介

鹿士义,男,1962年生,北京大学对外汉语教育学院副教授,中国科学院心理研究所博士后,研究方向为语言测试、语言习得。

王二平,男,1953年生,中国科学院心理研究所研究员,研究方向为组织心理学等。

(通讯地址:鹿士义 100871 北京大学对外汉语教育学院商务汉语考试研发办)